

Plagiarism in student papers

Prevalence estimation using special techniques for sensitive questions

Ben Jann

ETH Zurich, jannb@ethz.ch

Konstanz, October 26, 2009

Outline

- Introduction
 - Plagiarism
 - Asking Sensitive Questions
- Results from Three Studies
 - Study A: Randomized Response Technique
 - Study B: Item Count Technique
 - Study C: The Crosswise Model
- Conclusions

Plagiarism

- What is plagiarism?

Definition of the U.S. National Academy of Sciences

“Appropriation of another person’s ideas, processes, results, or words without giving appropriate credit, including those obtained through confidential review of others’ research proposals and manuscripts”

- In the age of the Internet, Wikipedia, etc. Universities increasingly begin to worry about plagiarism in student papers and homework assignments.

Plagiarism

- Universities usually provide leaflets to lecturers and students about definitions and consequences of plagiarism:

Plagiat: Hinweise zum Umgang mit Plagiaten

In der jüngsten Vergangenheit haben sich Fälle von Plagiaten bei Seminar- und Abschlussarbeiten sowie Hausarbeiten gehäuft. Unter Plagiat versteht man die wortwörtliche **Übernahme von Texten – sowohl aus dem Internet als auch aus Printmedien – ohne deren Kenntlichmachung als Quelle (Zitat)**. Dabei sehen die Regeln guter wissenschaftlicher Praxis generell vor, Gedanken Dritter durch Zitate (indirekt oder direkt) kenntlich zu machen. Das bedeutet: Quellen sind kenntlich zu machen und wörtlich übernommene Passagen mit Quellenangaben zu zitieren. Die „Richtlinien zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten an der Universität Konstanz“ finden Sie in den Amtlichen Bekanntmachungen der Universität Konstanz Nr. 8/98 vom 08.12.1998.

Link: http://www.uni-konstanz.de/struktur/org/org_files/richtlinien_wiss_praxis.pdf

Bitte beachten Sie, dass Plagiate **immer als Täuschungsversuch** angesehen werden und stets prüfungsrechtlich geahndet werden. Die betreffende Prüfungsleistung gilt in diesem Fall als „**nicht ausreichend**“, also **nicht bestanden**. Darüber hinaus können in schwerwiegenden Fällen weitergehende Sanktionen bis zum **Ausschluss von der Wiederholungsprüfung** und den damit verbundenen **Verlust des Prüfungsanspruchs in dem Studiengang** ausgesprochen werden, was die **Exmatrikulation** zur Folge hat. Die Einzelheiten richten sich nach den jeweiligen Prüfungsordnungen.

Link: <http://www.uni-konstanz.de/studium/index.php?cont=pruefung&lang=de>

<http://www.uni-konstanz.de/studium/pdf/pruefinfo/HinweiseUmgangmitPlagiaten.pdf>

Plagiarism

- “Degree” of plagiarism:
 - ▶ Partial plagiarism: e.g. include a text passage from someone else’s work without marking it as a citation
 - ▶ Severe plagiarism: e.g. have someone else write most of the paper or hand in a someone else’s paper
- It might be important for Universities to know something about how frequent different forms of plagiarism occur.
- How can prevalence estimates of plagiarism be obtained?

Plagiarism

- Approaches:
 - ▶ Official number of students found guilty.
 - ▶ Systematic inspection of a sample of student papers.
 - ▶ Prevalence estimation using hypothetical scenarios/vignettes (“How would you behave in this situation”?).
 - ▶ Prevalence estimation based on self-reported behavior (“Did you ever ...?”).
- Will the students be willing to admit that they are guilty of plagiarism if we ask them? What are the techniques we can use to gain valid answers to such questions?
- Some examples ...

Asking the Embarrassing Question

BY ALLEN H. BARTON

University of Chicago

THE POLLSTER's greatest ingenuity has been devoted to finding ways to ask embarrassing questions in non-embarrassing ways. We give here examples of a number of these techniques, as applied to the question, "Did you kill your wife?"

[The Public Opinion Quarterly 22/1958: 67-68]

1. The Casual Approach:

“Do you happen to have murdered your wife?”

2. The Numbered Card:

Would you please read off the number on this card which corresponds to what became of your wife?” (HAND CARD TO RESPONDENT)

1. Natural death

2. I killed her

3. Other (What?)

(GET CARD BACK FROM RESPONDENT BEFORE PROCEEDING!)

3. The Everybody Approach:

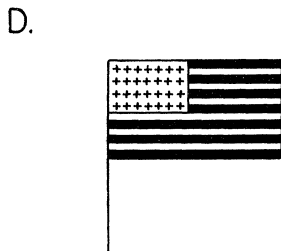
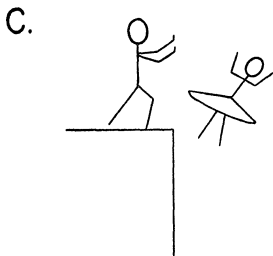
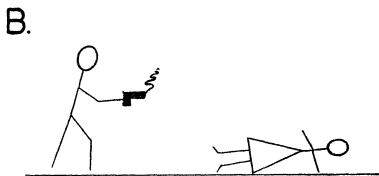
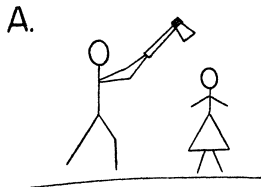
“As you know, many people have been killing their wives these days. Do you happened to have killed yours?”

4. The “Other people” Approach:

- (a) “Do you know any people who have murdered their wives?”
- (b) “How about yourself?”

5. The Sealed Ballot Technique:

In this version you explain that the survey respects people’s right to anonymity in respect to their marital relations, and that they themselves are to fill out the answer to the question, seal it in an envelope, and drop it in a box conspicuously labelled “Sealed Ballot Box” carried by the interviewer.



6. The Projective Technique:

“What thoughts come to mind as you look at the following pictures?”

(Note: The relevant responses will be evinced by picture D.)

7. The Kinsey Technique:

Stare firmly into respondent's eyes and ask in simple, clearcut language such as that to which the respondent is accustomed, and with an air of assuming that everyone has done everything,

“Did you ever kill your wife?”

8. Putting the question at the end of the interview.

Asking Sensitive Questions

- Gaining valid answers to sensitive questions is difficult. People may underreport sensitive behavior (while overreporting socially desirable behaviors).
- Various techniques have been developed to guarantee anonymity and minimize the respondent's feelings of jeopardy, so that more honest answers can be expected.
- We will now look at three such techniques:
 - ▶ The randomized response technique (RRT)
 - ▶ The item count technique (ICT/UCT)
 - ▶ The crosswise model

The Randomized Response Technique (RRT)

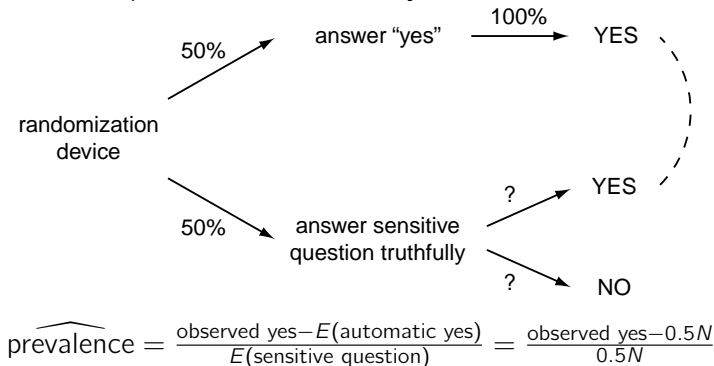
(Warner 1965; also see, e.g., Fox and Tracy 1986)

- Basic idea: **anonymity through randomization**.
- Depending on the outcome of a randomization device (e.g. roll a dice), the respondent has to answer the sensitive question or give an automatic “yes” or “no” answer (or answer an unthreatening question of which the distribution is known).
- Since only the respondent knows the outcome of the randomization device, a “yes” answer cannot be interpreted as an admission of guilt.
- However, the proportion of the sample that has engaged in the behavior of interest can be calculated with knowledge of the properties of the randomizing device.

The Randomized Response Technique (RRT)

(Warner 1965; also see, e.g., Fox and Tracy 1986)

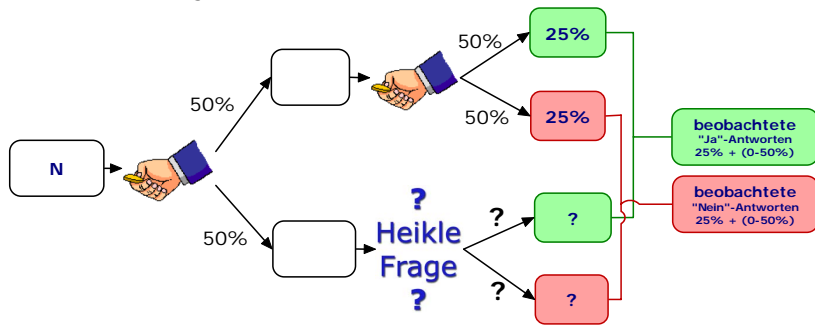
- Example (forced-response design): Toss a coin and, if heads, answer the sensitive question, else answer “yes”.



- **Critical assumption:** Respondents closely follow the instructions.


Study A: Using RRT to Measure Plagiarism

- Web-Survey among ETH students in 2005
- Response rate: 33 Percent
- Research team: Elisabeth Coutts, Andreas Diekmann, Georg Böcherer, Stefan Senn, Philipp Stadelmann, Diego Stutzer
- Used RRT-design:



Study A: Using RRT to Measure Plagiarism



Nimm bitte eine Münze  zur Hand und führe einen Münzwurf durch. Beantworte gemäss dem Ergebnis die entsprechende Frage:

Hast du Kopf geworfen, dann beantworte bitte die folgende Frage:

Hast du in einer dieser Arbeiten (Semester-, Bachelor-, Master- oder Diplomarbeit) schon einmal bewusst ein Zitat nicht gekennzeichnet?

Hast du Zahl geworfen, dann beantworte bitte die folgende Frage:

Bitte nimm **nochmals** die Münze  zur Hand und führe einen **Münzwurf** durch. Ist das Ergebnis 'Kopf' so beantworte die Frage mit Ja. Im anderen Fall beantworte die Frage mit Nein.

☐ Ja ☐ Nein

Study A: Using RRT to Measure Plagiarism

- Results:

<i>Geschätzter Anteil Plagiate bei...</i>	<i>Versuchsbedingung</i>		
	<i>direkte Befragung</i>	<i>RRT1</i>	<i>RRT2</i>
... einer Diplom-, Master-, Bachelor- oder Semesterarbeit	12.0%	2.0%	5.6%
... einer anderen Arbeit	19.4%	18.9%	16.2%

Study A: Using RRT to Measure Plagiarism

- Explanations:
 - ▶ difficulties understanding RRT, no trust in RRT
 - ▶ Web-surveys already anonymous enough?
 - ▶ False “no” bias: Respondents who did not commit plagiarism are reluctant to give a “yes” answer to the non-sensitive question.
- Approaches to deal with the false “no” bias
 - ▶ directly approach the problem using specific instructions
 - ▶ apply methods to detect cheaters and correct the RRT estimates
 - ▶ use alternative methods that are not (or less) affected by the false “no” bias

The Item Count Technique (ICT)

(see, e.g., Dalton et al. 1994, Raghavarao and Federer 1979)

- Given a list of statements, respondents report how many of them are true, but not which ones. For some respondents the list contains the sensitive item, for others not (randomized).
- Example: “How many of the following statements apply to you?”

Group A (short list)	Group B (long list)
I have a cat.	I have a cat.
I have blue eyes.	I have blue eyes.
I like country music.	I like country music.
	I use drugs.

- Prevalence estimate = mean difference
- Advantage: Requires no randomization device.

Study B: Using ICT to Measure Plagiarism

- Web-Survey among students of the University of Konstanz, Summer 2009
- Response rate: 23.7 Percent
- Research team: Ben Jann and Philipp Stirnemann (thanks to Thomas Hinz, Katrin Auspurg, and Pascal Gienger from the University of Konstanz for supporting the project)
- Questions:

Haben Sie beim Schreiben einer Hausarbeit (z.B. Seminararbeit, Semesterarbeit, Abschlussarbeit, etc.) schon einmal bewusst eine Textpassage aus einem fremden Werk übernommen, ohne diese als Zitat zu kennzeichnen?

- ☐ ja
- ☐ nein

Haben Sie schon einmal einen Grossteil einer Arbeit durch eine andere Person schreiben lassen oder eine fremde Arbeit (z.B. von www.hausarbeiten.de) als Ihre eigene ausgegeben?

- ☐ ja
- ☐ nein

Study B: Using ICT to Measure Plagiarism

Nachfolgend finden Sie vier Gruppen mit verschiedenen Aussagen. Zwei der Gruppen enthalten je eine Aussage, zu der man vielleicht nur ungern Auskunft gibt.

Zählen Sie deshalb bitte für jede Gruppe nur, wie viele der Aussagen Sie bejahen würden. Diese Zahl geben Sie dann für die entsprechende Gruppe an. Wenn zum Beispiel in einer Gruppe mit insgesamt fünf Aussagen drei auf Sie zutreffen, geben Sie für diese Gruppe als Antwort "3" an.

Diese Befragungsmethode garantiert Ihre Anonymität, da für uns nicht ersichtlich ist, welche der einzelnen Aussagen auf Sie zutreffen. Mit Hilfe der Wahrscheinlichkeitsrechnung ist es uns aber möglich, eine Häufigkeit für die Gesamtheit aller Befragten zu berechnen.

Gruppe 1:

- Ich bin ein sehr spontaner Mensch und manchmal auch ein bisschen chaotisch.
- Die Wahl der Uni fiel mir leicht, da ich mich einfach den Entscheidungen meiner Freunde anschloss.
- Ich spiele regelmässig Schach.
- Ich bin meistens sehr pünktlich.
- Beim Schreiben einer Hausarbeit (z.B. Seminararbeit, Semesterarbeit, Abschlussarbeit, etc.) habe ich schon einmal bewusst eine Textpassage aus einem fremden Werk übernommen, ohne diese als Zitat zu kennzeichnen.

Anzahl Aussagen, die Sie in dieser Gruppe mit "Ja" beantworten würden:

Study B: Using ICT to Measure Plagiarism

- Results: plagiarism prevalence estimates

	direct question (400)	ICT 1 (858)	ICT 2 (855)
partial plagiarism	8.0% (1.4%)	9.1% (5.3%)	10.4% (6.4%)
full plagiarism	2.0% (0.7%)	-6.8% (5.9%)	-0.2% (6.8%)

(standard errors in parentheses)

Study B: Using ICT to Measure Plagiarism

- By the way: Are students aware of the University's guidelines with respect to plagiarism?

An den meisten Universitäten gibt es klare Richtlinien zum Umgang mit Plagiaten. Sind Ihnen die entsprechenden Hinweise der Universität Konstanz zum diesem Thema bekannt?

- ☐ ja
- ☐ nein
- ☐ weiss nicht

Sind Sie sich ungefähr im Klaren, wann und wie man sich bei einer schriftlichen Arbeit eines Plagiats schuldig macht?

- ☐ ja
- ☐ nein
- ☐ hab ich mir noch nie überlegt

Sind Sie sich der Konsequenzen eines solchen Verstosses bewusst?

- ☐ ja
- ☐ nein
- ☐ hab ich mir noch nie überlegt

Study B: Using ICT to Measure Plagiarism

- By the way: Are students aware of the University's guidelines with respect to plagiarism?
- Results:
 - ▶ Aware of guidelines: 76.7% yes, 23.3% no/don't know
 - ▶ Clear what plagiarism is: 87.4% yes, 16.6% no/did not think about
 - ▶ Aware of consequences: 82.0% yes, 18.0% no/did not think about

The Crosswise Model

(Yu, Tian, and Tang 2007)

- Very simply idea: Ask a sensitive question and a non-sensitive question and let the respondent indicate ...
 - ▶ A: whether the answer is “yes” to both questions or “no” to both questions
 - ▶ B: whether the answer is “yes” to one questions and “no” to the other

		Question 1	
		no	yes
Question 2	no	A	B
	yes	B	A

- In either case, the researcher does not know whether the sensitive item applies to a specific respondent.
- The prevalence for the non-sensitive item must be known (and unequal to 0.5).

Study C: Using the Crosswise Model to Measure Plagiarism

- Classroom survey (written questionnaire) at different Universities (ETH Zurich, University Leipzig, LMU Munich), Spring/Summer 2009
- Total sample size approx. 500.
- Research team: Ben Jann, Julia Jerke, Ivar Krumpal (thanks to Norman Braun and Jochen Groß from LMU Munich for their support).

Study C: Using the Crosswise Model to Measure Plagiarism

Block 1

1. Frage: *Hat Ihre Mutter in den Monaten Januar, Februar oder März Geburtstag?*
2. Frage: *Haben Sie beim Schreiben einer Hausarbeit (z.B. Seminararbeit, Semesterarbeit, Abschlussarbeit, etc.) schon einmal bewusst eine Textpassage aus einem fremden Werk übernommen, ohne diese als Zitat zu kennzeichnen?*

Wie lauten Ihre Antworten auf die beiden Fragen?

- ☐ (A) auf beide Fragen *Nein* oder auf beide Fragen *Ja*
- ☐ (B) auf genau eine der beiden Fragen *Ja* und auf die andere *Nein*

Study C: Using the Crosswise Model to Measure Plagiarism

Results: plagiarism prevalence estimates

	direct (96)	crosswise (310)	ETH (54)	LMU (65)	Leipzig (191)
partial plagiarism	7.3% (2.7%)	22.3% (5.5%)	35.2% (13.6%)	8.5% (11.4%)	23.3% (7.0%)
full plagiarism	1.0% (1.0%)	1.6% (5.0%)	9.3% (12.5%)	-0.8% (10.8%)	0.3% (6.3%)

(standard errors in parentheses)

Conclusions

- Validity of estimates obtained using RRT is questionable (difficult to understand, the false “no” bias).
- There are methods that are easier to understand and give the respondent less incentive to deviate from the procedure.
- Results for these alternative methods are mixed.
 - ▶ ICT: Slightly higher prevalence estimates obtained for partial plagiarism compared to direct questioning. However, method failed for full plagiarism.
 - ▶ Crosswise model: Clearly increased prevalence estimate for partial plagiarism. No difference for full plagiarism.